Lecture 7: Structural Breaks, Dummy Variables and Multicollinearity

Prepared for ECON 5033

Eric S. Lin

Department of Economics, National Tsing Hua University

November 12, 2024

Outline

- Before moving to Generalized least square (GLS), we would learn some miscellaneous topics in multiple regression models.
- These topics include:
  - Testing for structural break
    - parameters may not be constant over time/units
  - Dummy variables
    - *regressors* or regressand taking only 0 or 1
  - Multicollinearity problem
    - correlations between regressors are too high

Motivation

- For the classical regression model, the implied assumption is that the regression coefficients are constant over time (over observations).
- Suppose that the MPC in consumption function changes at some point due to the new government policy, that is,

$$C_t = \beta_{11} + \beta_{12} Y_t + \varepsilon_t \quad \text{for } t = 1, ..., t_1$$
$$C_t = \beta_{21} + \beta_{22} Y_t + \varepsilon_t \quad \text{for } t = t_1, ..., T.$$

- One can do parameter constancy tests.
- According to the test of parameter constancy on intercept, slopes or both, we could classify the test for structural changes into three types.
  - Intercept and slopes differ across the two samples
  - Only intercept differs across the two samples
  - Only slopes differ across the two samples

Variable Intercept and Slopes

- We specify the following model with a structural break.

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \tag{1}$$

  where X_1 and X_2 indicate the $n_1 \times k$ and $n_2 \times k$ matrices. β_1 and β_2 are both $k \times 1$ vectors.

- The standard model with k regression coefficients is in fact a restricted model by imposing the following restriction.

$$R_{k\times 2k}\beta_{2k\times 1} = \begin{bmatrix} I_k & -I_k \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = \beta_1 - \beta_2 = 0_k.$$

- We've seen how to test the hypothesis $R\beta = 0$ in previous lectures.

$$F = \frac{\left(e^{*\prime}e^{*} - e'e\right)/k}{e'e/(n-2k)} \sim F\left(k, n-2k\right).$$

Variable Intercept Only

- We need to write down the unrestricted model (under structural change):

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \iota_{n_1} & 0 & X_{12} \\ 0 & \iota_{n_2} & X_{22} \end{bmatrix} \begin{bmatrix} \beta_{11} \\ \beta_{12} \\ \beta_2 \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}.$$

- The restriction (without structural change) is $R\beta = 0$.
- The test statistic is

$$F = \frac{\left(e^{*\prime}e^{*} - e'e\right)/1}{e'e/(n-(k+1))} \sim F\left(1, n-k-1\right).$$

Variable Slopes Only

- The unrestricted model is given by

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \iota_{n_1} & X_{12} & 0 \\ \iota_{n_2} & 0 & X_{22} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_{12} \\ \beta_{22} \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}.$$

- What does the restriction look like?
- The test statistic is

$$F = \frac{\left(e^{*\prime}e^{*} - e'e\right) / (k-1)}{e'e/(n-(2k-1))} \sim F\left(k-1, n-2k+1\right).$$

Chow Test

- The F test for testing the parameter constancy is usually referred to as Chow test (Chow, 1960)
- Gregory Chow is Professor of Economics at Princeton University.
- Gregory C. Chow (1960): "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica* 28 (3): 591–605
- There are **10,316** citations accumulated on Nov. 11, 2024
- Need to know the break point in advance. Can adjust the Chow test by testing for all possible breaks. The test statistic is the maximum of all F-statistics but its distribution is nonstandard. See Stock and Watson (2003).

Gregory Chow with Ben Bernanke

Test for Structural Break – A Special Case

- Consider the case of $n_2 < k$ for testing all the regression coefficients are different over the two sub-samples. Problem?
  - The model cannot be estimated in the second sub-period
  - The unrestricted estimates cannot be obtained by separate OLS regressions in each sub-sample
  - The restricted RSS can be computed as usual, $e^{*\prime} e^*$
  - Compute the unrestricted model $e_1' e_1$ as the unrestricted RSS
  - The second sub-sample contributes zero to the RSS
  - Fisher (1970) suggested to test the null hypothesis using,

$$F = \frac{\left(e^{*\prime} e^* - e_1' e_1\right) / n_2}{e_1' e_1 / (n_1 - k)} \sim F\left(n_2, n_1 - k\right).$$

Test for Structural Break – Heteroskedasticity

- A standard Chow test is based on the assumption that the variances of the error terms are constant.
- If there is heteroskedasticity for the restricted model, i.e., $\text{var}[\varepsilon_1] \neq \text{var}[\varepsilon_2]$.
- F type statistics could not be used anymore.
- Suppose we could estimate coefficients in two sub-samples, $b_1 = (X_1'X_1)^{-1} X_1'Y_1$ and $b_2 = (X_2'X_2)^{-1} X_2'Y_2$.
- Suppose under the null of no structural break we have that, $b_1 \xrightarrow{d} \mathcal{N}(\beta, V_1)$ and $b_2 \xrightarrow{d} \mathcal{N}(\beta, V_2)$.
- CMT suggests that under the null, $b_1 - b_2 \xrightarrow{d} \mathcal{N}(0, V_1 + V_2)$. And one could form a Wald type test.

Motivation

- A dummy variable is an indicator variable for whether a variable takes on a particular number or belongs to a particular category.
- For example, the dummy variable D_i^F agrees

$$
\begin{aligned}
D_i^F &= 1, \quad \text{if the } i\text{th observation is a female} \\
D_i^F &= 0, \quad \text{if the } i\text{th observation is a man.}
\end{aligned} \tag{2}
$$

- The group with $D_i^F = 0$ is referred to as the base, benchmark, reference, comparison, or control group.
- Dummy variable a.k.a. dichotomous, binary, discrete variables, categorical variables
- Dummy in regressors for today
- Dummy in regressand later on

Test for Structural Break in Terms of Dummy Variables

- We've learned the tests for structural break. It could be reformulated by the dummy variables.
- The model considering "variable intercept" is

$$Y_i = \alpha + \gamma D_i + \beta X_i + \varepsilon_i,$$

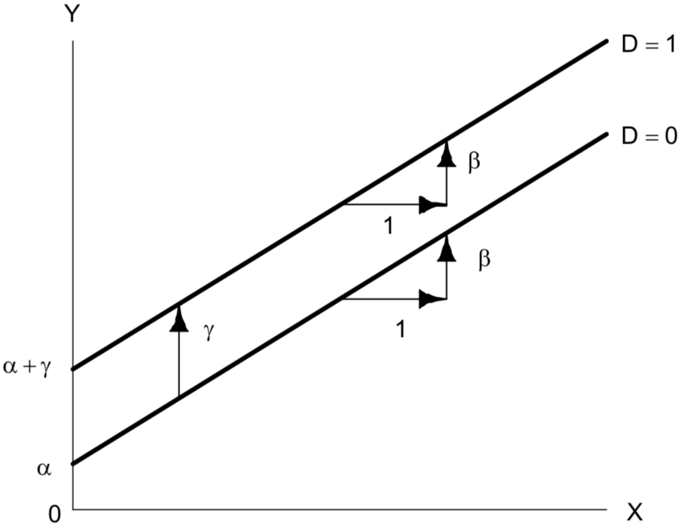
where $D_i = 0$ if $i \leq n_1$ and $D_i = 1$ if $i > n_1$.

- The model considering "both" is

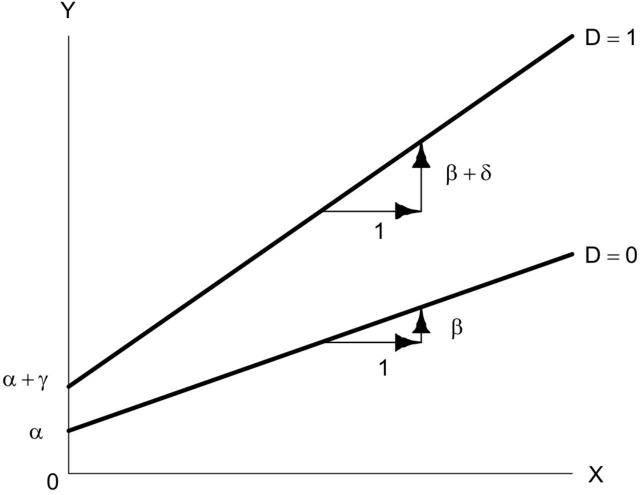
$$Y_i = \alpha + \gamma D_i + \beta X_i + \delta X_i D_i + \varepsilon_i,$$

where $D_i = 0$ if $i \leq n_1$ and $D_i = 1$ if $i > n_1$.

Graphical Illustration for the Case of Variable Intercept



Case of Variable Intercept & Slope



Marginal Effects?

- In the labor economics, we usually are interested in the wage differential between males and females.
- This could be done by setting the dummy variable like (2) in wage regression.

$$\ln(\text{Wage}_i) = \beta_1 + \beta_2 D_i^F + \beta_3 \exp_i + \beta_4 \text{edu}_i + \varepsilon_i.$$

- The coefficient β_2 is interpreted as the effect of being a female on wage relative being a male.
- Taking conditional expectation leads to

$$\begin{aligned} E[\ln(\text{Wage}_i)|D_i^F = 1, \exp_i, \text{edu}_i] &= \beta_1 + \beta_2 + \beta_3 \exp_i + \beta_4 \text{edu}_i \\ E[\ln(\text{Wage}_i)|D_i^F = 0, \exp_i, \text{edu}_i] &= \beta_1 + \beta_3 \exp_i + \beta_4 \text{edu}_i. \end{aligned}$$

- Interpret β_2? Geometrically?

Change of the Reference Group

- Now we re-define,

$$
\begin{aligned}
D_i^M &= 1, \quad \text{if the } i\text{th observation is a male} \\
D_i^M &= 0, \quad \text{if the } i\text{th observation is a female.}
\end{aligned}
\tag{3}
$$

- Run the following regression

$$\ln(\text{Wage}_i) = \beta_6 + \beta_5 D_i^M + \beta_3 \exp_i + \beta_4 \text{edu}_i + \varepsilon_i.$$

- Interpret β_5?
- Relationship between β_5 and β_2?

Dummy Variable Trap

- One may try to fit the following regression using two dummies,

$$\ln(\text{Wage}_i) = \beta_7 + \beta_8 D_i^M + \beta_9 D_i^F + \beta_3 \exp_i + \beta_4 \text{edu}_i + \varepsilon_i.$$

- In this case, we would get into trouble. (why?)
- Dummy Variable Trap!
- General rule for setting dummy variables: for a qualitative variable with J categories, one might include $J-1$ dummies in the regression model.

Three Groups

Example

Suppose there are three classes: Asian, American, and African American. Consider the following,

$$D_i^{AS} = \begin{cases} 1 & \text{if the } i\text{th observation is Asian} \\ 0 & \text{if the } i\text{th observation is not} \end{cases}$$

$$D_i^{AM} = \begin{cases} 1 & \text{if the } i\text{th observation is American} \\ 0 & \text{if the } i\text{th observation is not} \end{cases}.$$

We know that $D_i^{AS} = D_i^{AM} = 0$ will represent African American. The wage equation becomes,

$$\ln(\text{Wage}_i) = \beta_1 + \beta_2 D_i^F + \beta_3 \exp_i + \beta_4 \text{edu}_i + \beta_5 D_i^{AS} + \beta_6 D_i^{AM} + \varepsilon_i.$$

Scale and Intercept

- You may wonder why we should set the value of dummy to be one. Of course, you could set it as 2, 5 or 100. (why?)
- If there is no intercept term in your model, you would be allowed to incorporate J dummies in your regression model such as, $(J = 2)$

$$\ln(\text{Wage}_i) = \beta_8 D_i^M + \beta_9 D_i^F + \beta_3 \exp_i + \beta_4 \text{edu}_i + \varepsilon_i.$$

Seasonal Adjustment in FWL

Example

Consider the model with seasonal component.

$$\begin{aligned} Y_i &= \alpha_1 S_i^1 + \alpha_2 S_i^2 + \alpha_3 S_i^3 + \alpha_4 S_i^4 + X_{2i}\beta_2 + \varepsilon_i \\ &= X_{1i}\beta_1 + X_{2i}\beta_2 + \varepsilon_i, \end{aligned}$$

where S_i^j is the jth seasonal dummy, $j = 1, 2, 3,$ and 4. According to FWL theorem, $M_1 Y$ is a form of seasonal adjustment or called de-seasonalized. Furthermore, one could get the coefficient β_2 by running regression of de-seasonalized Y on de-seasonalized X_2.

Wage Differential

Example

Assume that the simple wage equation is

$$Y_i = \beta + \delta D_i^F + \varepsilon_i,$$

where $D_i^F = 1$, if the ith observation is a female. We could get the OLS estimators for β and δ.

$$\hat{\beta} = \frac{\sum_{i=1}^{n} Y_i \left(1 - D_i^F\right)}{n - \sum_{i=1}^{n} D_i^F}$$
$$\hat{\delta} + \hat{\beta} = \frac{\sum_{i=1}^{n} Y_i D_i^F}{\sum_{i=1}^{n} D_i^F}.$$

We could interpret $\hat{\beta}$ as the mean wage of men, $\hat{\delta} + \hat{\beta}$ as the mean wage of women and $\hat{\delta}$ being the wage differential.

Interaction Terms

- One can interact dummies to allow for combinations of categories. For instance,

$$\ln(\text{Wage}_i) = \beta_1 + \beta_2 D_i^F + \beta_3 \exp_i + \beta_4 \text{edu}_i + \beta_5 D_i^{AS} + \beta_6 D_i^{AM} + \beta_7 D_i^F \times D_i^{AS} + \varepsilon_i.$$

- The female wage differential is $\beta_2 + \beta_7 D_i^{AS}$.
- The Asian wage differential is $\beta_5 + \beta_7 D_i^F$.
- Another scenario is to interact dummy with other regressors.

$$\ln(\text{Wage}_i) = \beta_1 + \beta_2 D_i^F + \beta_3 \text{edu}_i + \beta_4 \text{edu}_i \times D_i^F + \varepsilon_i.$$

- In this case the effect of education on wage will depend on gender. How to test if the return to education is the same across gender?

Caveats on Program Evaluation

- A typical use of a dummy variable is when we are looking for a program effect or policy effect
- For example, we may have individuals that received job training, or welfare, etc

$$Y = \beta_1 + \beta_2 T_i + \beta_3 ... + \varepsilon_i$$

- We need to remember that usually individuals choose whether to participate in a program, which may lead to a self- selection problem
- If we can control for everything that is correlated with both participation and the outcome of interest then it's not a problem
- Often, though, there are unobservables that are correlated with participation – the estimate of the program effect is biased, and we don't want to set policy based on it

Motivation

- We say if the rank of regressor matrix X is less than k, there is multicollinearity problem.
- More formally, if $\text{rank}(X) < k$, the matrix $X'X$ will be singular in nature.
- The formula $\hat{\beta} = (X'X)^{-1} X'Y$ does not work in this situation.
- The conditional mean is $\text{E}[Y|X] = X\beta$. Given multicollinearity, one could have a $k \times 1$ nonzero vector η such that $X\eta = 0$.
- Conditional mean is still satisfied if we add $X\eta$, i.e., $\text{E}[Y|X] = X\beta + X\eta = X\gamma$. This means the regression coefficient β could not be identified if there exists multicollinearity.

Near Multicollinearity

- Multicollinearity could occur with inappropriate use of dummy variables – the dummy variable trap.
  - i.e. one of the regressor is exactly proportional to another regressor or is a linear combination of several other regressors.
  - perfect multicollinearity
- This could be solved by appropriate changes in the model.
- People often talk about "near" multicollinearity. It means that the regressors are fairly closely related so that although one could obtain estimates one cannot obtain precise estimates.
- Intuitively, when there are strong linear relations between the regressors, it is difficult to determine the separate influence of the regressors on the dependent variable.

A Working Example

- Let's consider the partitioned model,

$$Y_i = X_{1i}'\beta_1 + x_{ki}\beta_k + \varepsilon_i,$$

where x_{ki} and β_k are scalars.

- We would focus on the β_k coefficient and its variance, $\text{var}[\hat{\beta}_k] = [\sigma^2 (X'X)^{-1}]_{kk}$.
- Note

$$X'X = \begin{bmatrix} X_1'X_1 & X_1'x_k \\ x_k'X_1 & x_k'x_k \end{bmatrix}$$

and by partitioned inverse formula,

$$[(X'X)^{-1}]_{kk} = \frac{1}{x_k' M_1 x_k}.$$

A Working Example

- Consider the model of regressing x_k on X_1
  - What is RSS? What is R_k^2?

$$R_k^2 = 1 - \frac{x_k' M_1 x_k}{x_k' M_\iota x_k}.$$

- Now express variance of $\hat{\beta}_k$ in terms of R_k^2:

$$\text{var}[\hat{\beta}_k] = \frac{\sigma^2}{(1 - R_k^2) x_k' M_\iota x_k}.$$

- The standard error is too high if
  1. σ^2 is high.
  2. x_k does not vary much.
  3. small sample size.
  4. R_k^2 approaches one.

Methods Detecting Collinearity

1. Overall F test suggests jointly significant but the individual t tests gets nothing significant.
2. Coefficients may have the wrong sign or implausible magnitude.
3. Small changes in the data produce wide swings in the parameter estimates.
4. Some computer packages report the variance inflation factor (VIF),

$$\text{VIF} = \frac{1}{(1 - R_k^2)},$$

   for each coefficient in a regression model as a diagnostic statistic.
5. One could also check the condition number (CN) of $X'X$.

$$CN = \sqrt{\lambda_{\max}(X'X)/\lambda_{\min}(X'X)}$$

Alleviate Collinearity

- There are several methods to alleviate collinearity problem.
  1. Increasing the sample size.
  2. Principle components: Omitted! See Gurmu, Rilstone and Stern (1999).
  3. Try to impose the restriction on the original regression model if a priori information is available. Doing restricted least square will reduce the variance.
  4. Do stepwise regression based on reported VIF.
  5. Report the correlation matrix for regressors and drop highly correlated variables
  6. Do ridge regression.

Ridge Regression

- Hoerl and Kennard (1970)
- The idea is to add a scalar matrix to $X'X$ to make it less singular.
- The modified estimator (or Ridge regression estimator) is

$$\dot{\beta} = \left(X'X + \lambda I_k\right)^{-1} X'Y,$$

- Note that

$$\begin{aligned} \mathrm{E}[\dot{\beta}] &= \left(X'X + \lambda I_k\right)^{-1} X'X\beta \neq \beta \\ \dot{\beta} &\xrightarrow{p} \beta \\ \mathrm{var}[\dot{\beta}] &= \sigma^2 \left(X'X + \lambda I_k\right)^{-1} X'X \left(X'X + \lambda I_k\right)^{-1} < \mathrm{var}[\hat{\beta}]. \end{aligned}$$

- In fact, $\dot{\beta} = \arg\min_{\beta}(Y - X\beta)'(Y - X\beta) + \lambda\beta'\beta$